Contextual Recommendations using Intention Mining on Process Traces

Ghazaleh Khodabandelou

Contextual Recommendations using Intention Mining on Process Traces

Doctoral paper

Ghazaleh Khodabandelou
Centre de Recherche en Informatique
University of Paris 1 Panthéon-Sorbonne
Paris, France

ABSTRACT— Nowadays, digital traces are omnipresent in Information System (IS). Companies track IS interactions to retrieve and compile information about actors. Researchers of various streams, within IT and beyond, focused on recording actor interactions with systems and the technical possibilities to identify record and store these interactions. Tracing functionality has appeared in almost all common computer applications. This PhD project will focus on the establishment of a trace-based system and propose recommendations to actors regarding to their context. The objective of this thesis is to study process traces to propose recommendations to the actors by identifying a set of generic processes adaptable to the current actors' context. Thus, any actor, expert or novice, will be able to use this knowledge that gives contextual clues to identify the potential steps he could perform.

Keywords— intention mining; recommendations; process traces

I. INTRODUCTION

The approach of tracing and analyzing activities in Information Technologies (IT) context was born when the need of discovering process models has emerged [2]. It then spread on the software development domain [12,26]. The approach proposed in [19] aims to discover process models by using well-structured process models (workflow models). This kind of process is rigid and not flexible in the way of acting or thinking. It is like a predefined automaton that performs the tasks assigned to it, in a computerized way. This approach is mostly based on Petri Nets model and sequences of tasks. The study of traces in a real situation will define typical situations where certain methods are preferred. Processes are traditionally considered as sequences of steps, such as in workflow or activity diagrams [19, 2]. However, in practice, processes are subject to methodological emerging choices, they are creative in nature while having common elements; there are also specific facets for a process from an actor to another, etc. Traces process modeling must cover all these aspects: monitor decisions in the development process, identify common parts between processes and highlight specific facets.

We believe in the crucial role of the intentional aspect to discover process models. The actions performed by IS actors are outcomes of their latent motivation and purpose. While sequences of activities in processes have creative and dynamic nature and include actors' intentions, it is very restrictive just to consider the traces of actors' interactions as activity-oriented sequences [19]. MAP [1] is an intentional process metamodel with a navigational structure that supports the dynamic selection of the next intention to be fulfilled and the appropriate strategy to fulfill it, whereas guidelines help in the operationalization of the selected intention. The Map formalism models processes according to actors intentions, and supports process variability by defining different strategies to achieve them. Thereby, confronted to a specific situation and a particular intention of an actor, the process model reveals the strategy to fulfill this intention, and the navigation through the process can be pursued. We believe combining the Map intentional process metamodel with the analysis of activities' traces will allow us to rebuild process models in a form that is more suited to deal with the engineers' ways of thinking and working.

The proposed study will first focus on traces, trace-based systems and their use with the Map process metamodel. We will develop and test a tool in a real case in the context of Humanities analytical methods processes. This step requires to identify different methods of analysis and to obtain data to study. Then, from observations, the study will identify recurring processes in different contexts of use (e.g. regarding the type of studied data, data quantity, actor profile or even performed analysis beforehand), this part should lead to a better guidance in the analysis they want to perform.

The reminder of this paper is organized as follow: section II illustrates the problematic of our project, the hypothesis and raised questions. Section III describes related works and their differences according to this project. In Section IV, we define the main terms that we use in this paper and during this PhD thesis and we draw an outline of the method for each point of the problematic. Section V presents the proposed method and the related methodologies, which will be used to overcome the problems. Section VI considers method validation of this PhD project in different dimensions and the case studies. Finally, in section VII we conclude this paper.

II. PROBLEM STATEMENT

Empirical researches show the Information Systems Engineering (ISE) methods are underused [32]. This could be the result of an insufficient guidance when using a method. We try through this research project, to provide the situational recommendations to improve the use of ISE methods in practice according to actors' profiles, the type of data, research areas, available components, etc. This guidance might improve the quality of IT systems and increase the productivity. The traces generated by actors bring back information about their profiles; therefore, we can guide them by giving adequate recommendations suitable to projects context and their requirements and intentions. The research question is:

Q. Can we guess the intentions of IS actors to provide adapted recommendations?

In this perspective, we make several hypotheses about this technique performance and reliability, i.e., providing adapted recommendations by discovering the intentions of IS actors:

H_0. *It is not possible to make the recommendations providing such information.*
H_1. *It is possible to find the intentions by using adequate methods.*
H_2. *It is possible to make the adapted recommendations using the right intentions.*
H_3. *This technique of recommendations is more accurate than other techniques.*
H_4. *It is possible to find the actors' intentions and to provide efficient recommendations by this technique.*

In H_0, we consider the possibility that it is not possible at all to provide the recommendations from the actors' intentions or this technique will not be more efficient than others will. In second hypothesis H_1, we consider that there is the possibility to find the right intentions, i.e. the intentions related to observed actors' activities, if we use an adequate technique. The next hypothesis H_2 means that there is a possibility of providing the recommendations adapted to the situation if we find the right intentions. We suppose in H_3 that the supposed technique of recommendations is more efficient than other existing techniques. Actually, recommendations techniques provide only simple guidance; moreover, they are not intention-based so they cannot have a global view of the model. Finally, in H_4 we consider that it is possible to provide recommendations by finding the intentions and that the proposed approach is the most efficient. Based on H_4 we raise some questions about its implementation: Q_1) How to retrieve the traces related to interactions of IS actors with the preservation of their meaning? Q_2) How to retrieve the sequences of activities from the traces and give them sense? Q_3) How to discover the intentional aspect of activities' sequences taking into account their variability and anticipate the most likely sequences of intentions and/or activities? Q_4) How to provide recommendations using the knowledge obtained from the previous steps?

The first challenge is to extract traces from a myriad of data sources, e.g. databases, files, messages logs, etc. There are two facets for traces recording and retrieving: *syntax* and *semantic*. The traces generated by actors are digital data that are meaningless when they are not recorded in their context, i.e., they should be collected and then recorded retaining both aspect of semantic and syntax. On the one hand, there are many kinds of trace-based tools conceived to extract and record IS traces, on the other hand, there are many criteria to choose one tool among all. This is a COTS (Commercial-Off-The-Shelf) problem, which deals with both of the requirements and the existent tools in market to find a compromise between them.

Once we have selected and implemented an adequate tool to collect the traces, we need to know about the signification of the traces. These traces contain the information about the actors' actions but other information is also incorporated (e.g. system messages, registry string messages). Therefore, the actors' sequences of activities should be isolated from these traces using a suitable method. On the other hand, the sequences of activities that have been carried out, according to various actors' intentions do not contain any information about the actors' intentions directly. We can only observe the sequence of activities recorded by the trace-based tool and there is no possibility to link automatically the intentions to the activities. Hence, we need a suitable method to distinguish these intentions.

Another aspect is to anticipate the actors' intentions and activities. If we know about future actions of actors, we know consequently about their future intentions or vice versa. This could be preventive to any dysfunction or malfunction of IS methods. Moreover, the size of the sequence of activities is variable regarding the related intentions; e.g., in a given time, when an actor is surfing on the Web, the activities sequence is shorter than when programming in a coding language. Thus, the size of sequences of activities depends on the nature of intentions and it is not a constant value. In order to fragment these sequences into intentions, we need the appropriate methods and algorithms to identify when a sequence is finished and when another one gets started. Once the sequence is fragmented into intentions, we need to determine which fragment corresponds to which intention. Hence, a pertinent and efficient method is necessary to overcome all of these points.

Once we obtain the knowledge about the sequences of activities and the related sequences of intentions and the future sequences of activities/intentions, we can guide the IS actors to enact more adequate method while remaining in their project perspective. Understanding how IT projects are handled and which methods are used or likely to be used, are promising to help organizations improve the quality of their products and increase their productivity. This delicate operation needs a scientific method and the expertise of experimented engineers. In section V, we will propose a method and methodologies to overcome all of the mentioned aspects.

III. RELATED WORKS

Hereafter, we describe three fields of researches that have some similarities with our researches.

A. *Process mining approaches*

The needs to analyze business processes motivate the emergence of the *process mining* discipline which main goal is to extract information from event log to discover process model. *Process mining* is based on data mining; however, according to [2], "existing data mining techniques are too data-centric to provide universal comprehension of the end-to-end process in organizations". Apart from discovering actual process model, process mining also aims to verify conformance and enhancement of processes (i.e. increasing the link between the real activities and the prescribed activities). Process mining techniques settle between data mining and machine learning [2]. In the context of process mining, event logs are generated through a workflow engine and follow *Petri nets* formalism to express process model using different algorithms. These algorithms diversify the different process mining approaches

[23], which aim to discover underlying processes from event log. Some of these are α-algorithm [2], Directed acyclic graphs [8], hierarchical clustering [17], genetic algorithm [13], Instance graphs [26], and Inductive workflow acquisition [9].

The majority of these algorithms require an event log that contains several execution traces for the same process instance [23]. For most of them, it is necessary knowing the process instance identifier, which illustrates the process instance for each event. Moreover, particular situation or errors could confuse the algorithms and destruct the results. Other information that is necessary for algorithms such as instance graphs and α-algorithm is a threshold as algorithm parameter. Additionally to these disadvantages, they only take into consideration sequence of actors' activities to discover process model, in another words, they are activity-oriented approaches whereas to discover the process model we rely on the intentional aspect of the processes, which is based on the Map process metamodel and we call it intention mining.

B. Goal-oriented approaches

There are some approaches of intentional process metamodel to formalize processes. The common aim of these goal-modeling approaches is modeling the processes according to the purpose of the actors/projects/organizations. We quote among them i* [31], KAOS [24], MAP [1].

KAOS proposes to specify the system and its environment by a requirements model as instances of a conceptual metamodel to support the goals, agents, alternatives, etc. It is based on a goals diagram where goals are related through AND/OR decomposition links. To define systems requirements this decomposition refines high-level goals identified by actors into thinner particle of goals. This refinement requires classifying goals according to their level of abstractions and links the same goals to the same level of abstraction. This approach supports variability and have a well-structured semantic but is less involved in intentional aspect of IS actors. Moreover, it has rigid task-decompositions, which makes the common goals for the tasks complex phenomena with an artificial complexity.

Another process modeling framework is i*, a modeling language developed to analyze IS and the organizations environments of organizations and to model the relationship between actors and theirs goals. This framework supports both agent-oriented and goal-oriented modeling. The agents are autonomous because of their uncontrollable and non-cognizable behavior and the fact that an actor reasons from its own point of view. The i* model claims to capture what, how and why a part of software is developed. It allows evaluating the functional or non-functional requirements of systems. However, this modeling language has an operational semantic for the tasks and does not support it for the goals and it is not designed to be a variable framework, i.e., any variability for goals. The KAOS differs from i* framework by using an ontology, thus that is not a social approach.

The lack of variability, rigid task-decompositions and operational semantic of tasks, encourage us to approach other kind of formalisms. Map modeling language [1] [5] is an intentional process metamodel that allows formalizing flexible processes. It supports variability for goals and offers the possibility to follow different strategies by focusing on the intentional aspect when running methodological processes. During its enactment, a process is not limited to linear activities; actors, according to their context, have a variety of choices to execute a task. Map models (instances of Map metamodel) guide the actor by proposing dynamic choices according to their intentions. They can be executed non-sequentially and enacted as long as the intention is not completely fulfilled. Thereby, map process models offer a better adaptability to the context of each actor.

C. Machine Learning approaches in process mining

Extracting the information from a given sequence of activities is crucial to know about the intentional aspect of processes. The machine learning techniques sound promising to extract underlying intentions. Classification is a basic task in machine learning and consists in separating intentions into different classes. Some approaches to classify traces in process mining context exist [4,21]; in [21] a *trace clustering* approach is used to overcome the problems of non-well-structured processes by classifying event logs in terms of cases. [22] uses a machine learning technique to learn models from data. These models are useful to provide insights into business processes. These approaches focus only on activities sequences and do not take into consideration the intentions behind these activities. Moreover, many classical techniques of classification such as SVM [33] or k-means cannot deal with the noise (incomplete or irrelevant data) and do not support the variability of data sequences - they only accept a predefined length of sequences whereas sequences of activities length is variable according to actors' purposes. As mentioned earlier, Machine Learning approaches are applied on the logs in an activity-oriented context. The meta-models based on activities are rigid and limited to a low-level of tasks execution; therefore, they miss a huge part of information (intentions) that allows having a overall view on the process. For these first reasons and the further ones (cf. Section V. Part C) Hidden Markov Models [29] seem to be promising to tackle with the Intention Mining' challenges.

D. Recommendation techniques

In recent years, some approaches have been proposed on the generation of recommendations as possible next steps, i.e., use event logs to guess which activity may follow a current activity. [6] is proposed in a context of flexible processes, i.e., declarative processes, based on historic information, i.e., traces. It shows that the process performance is higher with an appropriate guided selection of activities. Others are data-oriented approaches based on optimization technique [18].

IV. TERMINOLOGY

In this section, we clarify some terms that are totally dependent on the manner of designing system model.

A. Actors

Actors are active entities involved in IS that carry out independent actions. They can be humans, automated machines (hardware and software) or combinations of them. Their behaviors are not completely random but not completely predictable. According to [18], software actors are inherently autonomous and they have social behaviors to achieve goals, i.e.

they communicate, coordinate, and cooperate. Human actors can be categorized into several levels, i.e., an individual person, actors as a group of persons, combination of groups, etc. The IS tools can also be determined in various levels of acting, e.g. considering a group of tools as actor(s), each tool as an actor, combination of a tool(s) and human actor(s) as actor(s), etc. This model is inspired by what is expected from any system. In our context of intentional modeling language, we set the borders of the actors at an active individual human actor accompanied by his IS tool. This actor carries out the various strategies using IS to achieve the goals of software development project at intentional level.

B. Intentions

The behaviors of actors widely depend on their intentions but not all of these are detectable, thus, it is important to model the level of abstraction for intentions. We define the intentions as the motivations to achieve the goals of development project. The possible motivations that do not lead to the goals are considered as *non-intentional*. The intentions in a software development context could emanate from an individual human actor, a group of human actors, delegation of a task, and related intention, from actor(s) to another or to a group of actors, nested delegation of tasks (and related intentions) from actor(s) to an actor or the groups of actors, etc. Hence, the relationships between actors characterize the abstraction level of intentions. An intentional ontology phase is necessary to analyze the alternatives of the actors in each step. We limit the boundary of human actors' intentions to individual intention level.

C. Actions

The activities are the actual actors actions recorded by IS tool. They are carried out to achieve goals according to the actors' intentions. There are different kind of actions with respect to the actors' relationships, intentions and environments. There are the *intentional*, *non-intentional* and *accidental* actions. As we defined in previous section, in our context the relationship of actors are limited to one actor and its IS tool. Thus, there is no task delegation between actors and no *nested-intention* accordingly. Therefore, all kind of activities recorded are performed by an actor on his IS tool with his own intentions. This definition encompasses intentional actions of actors. However, several actions are the product of *non-intentional* motivations; we call them *non-intentional* actions. Third kinds of actions are *accidental* ones, which are probably produced by mishandling of keyboard, internal system tools interactions, errors of OS, errors of IDE, etc. We consider *non-intentional* and *accidental* actions as traces noise. According to the degree of noise robustness, intention-mining algorithms can mitigate these non-desirable actions and their influence on results.

V. METHOD AND METHODOLOGY

In order to overcome the problems mentioned in section II, we propose the following method:

a) Retrieving digital traces with an adequate tool and recording them while preserving their contextual aspect.
b) Extracting sequences of activities from the retrieved traces.
c) Discovering the intentional aspect of activities' sequences taking into account their variability and anticipate the likely sequences of intentions and/or activities.
d) Choosing the proper method to provide recommendations.

For each point of the previous section, we propose the following methodologies.

A. Systematic review of trace-based tools

In order to collect the activities traces, we need a suitable trace-based tool that conserves contextual aspect of traces, i.e., conserving the order of execution of activities, the profile of actors and projects in a temporal context. We studied in [30] some trace-based tools to find the adequate one that responds to the project requirements. We conducted this research by carrying out a phase of requirements analysis using a decision-making method. Thereby we defined a set of requirements and we searched for a tool that fitted those requirements. At this stage, we were confronted to a *decision making* problem to which we responded by adopting a particular method. This study showed that *Snare* [34] respond to 87,4 % of the defined requirements. Thus, we will use Snare to collect traces of activities and retrieve information. The process of requirements analysis was a challenge, because of the need to analyze actors' behaviors according to their intentions, the variety of event log sources and the involvement of various part of IS. Moreover, defining requirements becomes more complex with the increase of the research space. We confronted challenging aspects in a detailed study of many trace-based tools, to find tools with a high degree of satisfaction according to our requirements and in the same time free and open source. This step of the methodology allows us responding to the first proposed method, i.e. retrieving digital traces with an adequate tool and recording them while preserving their contextual aspect.

B. Transforming the traces of activities into sequences of activities

In order to preserve the semantic of traces, according to [16] we can extract directly from IS a qualitative trace (semantic aspect) from quantitative data (syntax aspect). One solution can be to attach or integrate a plug-in to the tools to give a posteriori meaning to the logs and define them as the sequence of activities. The event logs collected by Snare have various attribution fields such as time, date, actor ID, system name, event ID, source of audit and strings. The strings have a long message illustrating the actions performed by actors. The set of these actions are the sequences of activities and input of intention-mining algorithms. To isolate the actions in each event log to construct the sequence of activities, we define ontology of activities to give them a posteriori meaning.

C. Choosing the method of intention mining

1) Critics of selection method

At this stage, we need to select a method to analyze the traces of activities, which is the third step of the method. The traditional approaches of process mining mentioned in section III do not deal pertinently when traces have ad-hoc behavior and high degree of variability, e.g. having the different intentions in a given time, changing the intentions suddenly, non-intentional motivations, etc. This kind of algorithms has a good performance when they are applied on well-structured processes such as workflow processes [14], they are activity-oriented and do not consider the intentions, thus they have predictable behaviors. However, Intention approaches do not model a process only as a set of tasks as *Petri nets* try to do, but as a set of intentions which are the sources of the sequence of tasks. On the one hand, the traces of activities have a probabilistic nature thus; we need a probabilistic model that affords a high degree of flexibility. On the other hand, activities' sequences have a high degree of variability that depends on the nature of the intentions: for instance, when an actor's intention is to make a wire transfer on the Internet, the activities to achieve this task are clearly defined and have to be performed in a sequential manner. On the contrary, if his intention is to develop a software, the activities involved to perform the task are more diversified (establish a project roadmap, manage a team, development in an IDE, etc.). Consequently, the nature and the length of the activities sequence are different from one intention to another. Moreover, for two similar intentions, the length of the activities sequences might also be variable. Unfortunately, classical classification algorithms do not support well this variability as they give results with sequences of constant length - such a strong constraint does not suit our needs. Therefore, we need a method that can deal with variability of sequences of activities length. Using a probabilistic model allows: a) knowing about the nature of data; if data is an outcome of meaningful tasks or only accidental behaviors, b) considering the temporal aspect of data execution, c) modeling latent parts of observed data, and d) extracting the features and structure of both observed and latent data. [11] considers Markov models as versatile since it is a hybrid statistical and algorithmic approach.

Hidden Markov Models (HMMs) [29,25] are promising to cover all the points raised before, i.e., discover the intentional aspect of activities' sequences, taking into account their variability and their probabilistic nature and anticipate the likely set of intentions and/or activities. HMMs are a stochastic generalization of finite-state automaton that evaluates the transitions probability between states and distributions probability of observations in those states. HMMs are flexible since they can model the complex structure of temporal dependencies between states. Thus, this flexibility allows adapting the process models to the context in a dynamic way. [20] considers HMMs as robust to noise with a controllable complexity using probability threshold parameter.

2) Inference of process metamodel

HMMs are stochastic signal models that allow modeling observed sequence in terms of a finite number of hidden states. We try to model, in our framework, the intentions as the hidden states and the observations as activities' sequences. We will use HMMs in supervised and unsupervised approaches. In both, HMMs present several challenges, such as defining the topological structure of the model, understanding the best ways to estimate the HMMs parameters, evaluating the probability of a given activities' sequence, evaluating the most likely set of intentions associated to a given activities' sequence, etc. Additionally, in unsupervised approach HMMs, we have to deal with problems such as defining the number of intentions. We showed the result of supervised approach in [10]. Some approaches of process mining have used HMMs, for instance in [7], to evaluate the quality of discovered process models they map Petri nets to HMMs by assigning the transitions as states. [9] considers the states of HMMs as activity nodes.

We will extend our approach to unsupervised approach. Both supervised and unsupervised approaches allow discovering, conforming and enhancing process models. We plan to use the Baum-Welch algorithm [15,28] and inverse problem techniques.

D. Choosing proper methods to make the recommendations

Once the sequences of intentions are evaluated, we then propose recommendations using the Map process metamodel with the intentional aspect of actors' activities. Given a specific intention, the map model will propose a dynamic choice of strategies to realize it (dynamically, since actors may change their intentions during the process). Knowing the intentions of actors allows detecting any deviation from prescribed process model. Moreover, this knowledge helps guiding actors to make better decisions based on more adapted recommendations. Indeed, a link between intentions and activities sequences is necessary.

Data analysis techniques such as Principal Components Analysis (PCA), Factor Analysis (FA), Independent Components Analysis (ICA) and Discriminate Analysis (DA) are the main classical methods for analyzing and reducing high dimensional complex data, i.e., multivariate, heterogeneous, with great dimensionality, with missing data and observed at different sampling rates. The *trace indicator values* of a project (e.g. size, duration, cost, application type) and actors (e.g. role, age, gender, experience, expertise) are recorded in a trace [3]. These indicators concern the user who enacts the process, his expertise and experience regarding the project, the innovative nature of the project, etc.

VI. Validation

A. Methods Evaluation considered

Our purpose is to demonstrate that the proposed method corresponds to the use for which it is provided. Method evaluation consists in all the operations necessary to prove that the method is sufficiently accurate, reliable and to have confidence in the results. In order to verify the method performance in high scale, we will apply them on real data in different case studies. Another method of evaluation that we will use is controlled experience, which consists in applying the proposed method on real cases in companies. Controlled experience allows showing the feasibility of the method in real cases. It allows knowing a priori and/or a posteriori the use of the proposed method, model and tool, knowing the user practices, their expectations and needs, knowing the opinions of the users about concepts or applications, assessing the interest and satisfaction with a product (method, model, tool).

We can do controlled experience in quantitative step and qualitative step. Quantitative step aims at quantifying behaviors, expectations or needs within a population. That can be done by a sample survey, one can ask the experts of this field to criticize the method by a closed questionnaire during an interview and capturing the traces when using an IS and applying the theoretical method on. Qualitative techniques focus on users' meeting and interview semi-directive, etc.

B. *Validation of the method with case studies*

The first application of this research concerns a transdisciplinary project with the Humanities department of our university, with the analysis of scientific methods used in their fields. It will provide medium-term recommendations to Humanities researchers when using tools like statistical analysis ones. Scientific processes of Humanities data analysis are multiple and complex [27]. There are many procedures, combining various methods of analysis, which vary according to studied data or the situation of ongoing research. However, there is no repository to share these processes and the associated knowledge to reuse them. Therefore, there is a loss of knowledge related to different analysis processes. Researchers must reinvent the analysis processes, even if other researchers have previously done similar analyses.

The second application of this work concerns the domain of situational method engineering and is a joint research with the Slovenian research Laboratory of Data Technologies (LPT). Here, we apply traces and trace analysis to the engineering of situational methods to analyze the ways of working of actors during information systems engineering projects: the purpose of the project is to identify the deviation of the process traces from the theoretical process model that should have been used. Moreover, we hope that our approach will allow us to provide recommendations adapted to the context of actors during the enactment of engineering methods.

VII. CONCLUSION

In this doctoral paper, we described the methodology we will follow during this PhD project. Our goal is to answer the following question: given the under-use of ISE methods, can we guess the intentions of IS actors to provide adapted recommendations? We will follow a five steps methodology. First, we select a tool to retrieve digital traces of users while preserving their contextual aspect. Second, we extract sequences of users' activities from these traces. Third, we define a method to find the intentions hidden behind the users' activities. Fourth, according to the users' intentions, we develop a method to provide proper recommendations. Finally, we will validate the method with appropriate tools, and practical case studies.

REFERENCES

[1] C. Rolland, N. Prakash, and A. Benjamen, "A Multi-Model View of Process Modelling," Req. Eng., vol.4, N°.4. Springer-Verlag, 1999.

[2] W.M.P. Van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes," 1st Ed., 2011, pp. 184-352.

[3] C. Hug, R. Deneckere, and C. Salinesi, "Map-TBS: Map process enactment traces and analysis," RCIS, 2012, pp. 204-209.

[4] L. Maruster, "Machine Learning Approach to Understand Business Processes," PhD Thesis, Tech. Univ. Eindhoven, The Netherlands, 2006.

[5] C. Rolland, "A Comprehensive View of Process Engineering," CAiSE'98, LNCS 1413, Springer, Pisa, Italy, June 1998.

[6] H. Schonenberg, B. Weber and B.V., Dongen, "Supporting Flexible Processes through Recommendations Based on History," Springer, BPM, vol. 5240, pp. 51–66. 2008.

[7] A. Rozinat, M. Veloso, and W.M.P. van der Aalst, "Evaluating the quality of discovered process models," 2nd Int. Workshop on the Induction of Process Models, Antwerp, Belgium, 2008, pp. 45-52.

[8] R. Agrawal, D. Gunopulos, and F. Leymann, "Mining Process Models from Workflow Logs", EDBT, LNCS 1377, pp.469-483, Springer, 1998.

[9] J. Herbst, and D. Karagiannis, "Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models," Workshop on Database and Expert Systems Applications, 1998.

[10] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi, "Supervised Intentional Process Models Discovery using Hidden Markov Models," 7th IEEE RCIS conference, Paris, France, 2013.

[11] J. Cook, and A. Wolf, "Automating process discovery through event-data analysis," Int. Conf. on Soft. Eng., ACM Press, 1995, pp.73-82.

[12] F. Harmsen, S. Brinkkemper and J. L. Han Oei, "Situational method engineering for information system project approaches," Methods and Associated Tools for the Information Systems Life Cycle conference, 1994

[13] W. van der Aalst, A. Medeiros, and A. Weijters, "Genetic Process Mining," Applications and Theory of Petri Nets, LNCS 3536, 2005.

[14] W. van der Aalst, A. Weijters and L. Maruster, "Workow Mining: Discovering Process Models from Event Logs," IEEE Trans. on Knowledge and Data Eng, vol. 16, N° 9, 2004, pp. 1128-1142.

[15] T.K. Moon, "The expectation-maximization algorithm, "Signal processing magazine, IEEE, vol.13, N°6, 1996, pp. 47-60.

[16] J. Laflaquière, "Conception de système à base de traces numériques pour les environnements informatiques documentaires," PhD thesis, Université de Technologie de Troyes, 2009.

[17] G. Greco, A. Guzzo, and L. Pontieri, "Mining Hierarchies of Models: From Abstract Views to Concrete Specifications," BPM, 2005, pp. 32-47.

[18] I.Vanderfeesten, H. Reijers, W. van der Aalst, " Product-based workflow support. Information Systems," vol. 36, N°2, pp. 517-535, 2011.

[19] G. Grosz, C. Rolland, S. Schwer, C. Souveyet, V. Plihon, S. Si-Said, C. Ben ashour and C. Gnaho, "Modelling the requirements engineering process: An overview of the nature approach," RE'97,London, vol.2,1997.

[20] J. Cook and A. Wolf, "Discovering Models of Software Processes from Event-Based Data," ACM Trans. on Soft. Eng. and Methodology, vol.7, N° 3, 1998, pp. 215–249.

[21] S. Minseok, C.W. Günther, and W.M.P. van der Aalst, "Trace Clustering in Process Mining Business Process Management Workshops," LNBIP 2009, vol.17, pp. 109-120.

[22] F. Eibe and I. H.Witten, "Data Mining: Practical Machine Learning Tools and Techniques," 3rd Ed., Morgan Kaufmann Publishers Inc. San Francisco, USA, 2005.

[23] D. Ferreira, M. Zacarias, M. Malheiros and P. Ferreira, "Approaching Process Mining with Sequence Clustering: Experiments and findings," BPM 2007. LNCS, vol. 4714, Springer, Heidelberg, 2007, pp. 360–374.

[24] A. Dardenne, A. van Lamsweerde and S. Fickasu, "Goal-directed requirements acquisition," Sci. Comput. Program., vol. 20, N°. 1–2, 1993.

[25] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," IEEE, vol.77, N°.2, 1989, pp. 257-286.

[26] B. van Dongen, and W. van der Aalst, "Multi-Phase Process Mining: Building Instance Graphs," Int. Conf. on Conceptual Modeling, LNCS 3288, Springer, 2004, pp.362-376.

[27] F. Djindjian, H. Noizet, L.Costa, and F. Pouget, "Webmapping dans les Sciences historiques et archéologiques," Archeologia E Calcolatori, n°19, Firenze : All'insegna Del Giglio, 2009, pp. 237.

[28] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," Ann. Math. Statist., vol.41, N°1, 1970, pp. 164-171.

[29] B.H. Juang and L.R. Rabiner, "A Probabilistic Distance Measure for Hidden Markov Models," AT&T Technical Journal, vol.64, N°2, 1985, pp. 391-408.

[30] G. Khodabandelou, C. Hug, R. Deneckère, C. Salinesi, M. Bajec, E. kornyshova and M. Jankovic, "COTS Products to trace method enactment: review and selection", 21st ECIS, Utrecht, Netherland, 2013.

[31] E. Yu, Modelling Strategic Relationships for Process Reengineering, PhD thesis, University of Toronto, Dpt of Computer Science, 1995.

[32] I. Mirbel and V. de Rivieres, "Adapting Analysis and Design to Software Context: The Jecko Approach," OOIS, 2002.

[33] T. Joachims, "Text categorization with support vector machines," Technical report, LS VIII N° 23, University of Dortmund, 1997.

[34] http://intersectalliance.com/projects/index.html, event log analyzer, viewed October 2012.